UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/837,158	04/19/2001	Karen Mae Holland	ARC000018US1	8446

48146 7590 05/15/2006

MCGINN INTELLECTUAL PROPERTY LAW GROUP, PLLC
8321 OLD COURTHOUSE ROAD
SUITE 200
VIENNA, VA 22182-3817

EXAMINER	
BLACKWELL, JAMES H	

ART UNIT	PAPER NUMBER
2176	

DATE MAILED: 05/15/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/837,158	HOLLAND ET AL.	
	Examiner	Art Unit	
	James H. Blackwell	2176	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *04 January 2006*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-12 and 14-28* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-12,14-25,27 and 28* is/are rejected.

7)☒ Claim(s) *26* is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *09 April 2001* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

1. This Office Action is in response to Amendment received 02/21/2006 with an original priority date of **04/19/2001**.

2. Claims 1-12, and 14-28 are pending. Claims 1, 17, and 23 are independent claims.

3. Claims 25-28 are new claims.

Allowable Subject Matter

4. Claim 26 is objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

Claim Rejections - 35 USC § 103

5. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

6. Claims 1-2, 17, 23, 25, 27-28 and 14-24 are rejected under 35 U.S.C. 103(a) as being unpatentable over Lewak et al. (hereinafter Lewak, U.S. Patent No. 5,544,360 filed 02/03/1995) in view of Goldman et al. (hereinafter Goldman, "Knowledge Discovery in an Earthquake Text Database: Correlation Between Significant Earthquakes and the Time of Day", Copyright 1997, IEEE, Pgs. 12-21).

In regard to independent Claim 1 (and similarly independent Claims 17, and 23), Lewak teaches *generating a dictionary of keywords in said text documents* in that the user, or an automated process (Col. 9, lines 50-55) analyzes each uncategorized file and can define categories (keywords) from those documents (Col. 8, lines 14-15; 61-65).

It is noted that the files taught by Lewak can contain files containing text along with other types of files (assuming that Fig. 1 represents files typical of those contemplated by the invention).

It is also noted that the categories (*keywords*) contained in a list that the user or a automated system that can be assigned can contain the *structured variables* as claimed (see Fig. 5, items listed in box containing item 52, specifically categories containing dates).

Lewak also teaches *forming categories of said text documents using said dictionary and an automated algorithm* in that the user, or an automated process can further group files containing similar sub-groupings together (Col. 9, lines 56-67; Col. 10, lines 1-10).

Lewak also teaches *counting occurrences of said structured variables, said categories, and combinations of said structured variables and said categories for said text documents* in that each category has an associated data structure record, which, among other things, stores how many files use that category along with linking and identifiers of the categories assigned to (Col. 5, lines 40-60). Given that such data is kept implies that such occurrences were tabulated.

Lewak fails to teach *identifying a relationship between a structured variable of said structured variables and text documents included in a category of said categories based on a probability of occurrence of a combination of said structured variables and said category*. However, Goldman teaches creating a dictionary of words that are present in the entire collection of texts (excluding stop words such as and, the, and or). At the end of this processing, a document is now a vector of normalized frequencies of words in the global dictionary. Thus, we now have a collection of vectors that can be measured and compared (Pg. 12, Sec. 2). A "database" of sorts (text-based) is then constructed with further filtering. Given that the goal of this data mining exercise is to look for relationships between Earthquakes in California (a sub-category of the Earthquake documents) and time of day (structure variable), some subject-specific words are isolated like location words (see Pg. 14, Tables 2-3). Figure 1 represents the frequency of occurrence of Earthquakes in California (category) v. time of day (structured variable). In other words, Golden teaches that those documents containing data for Earthquakes in California over a given number of years versus time of day are mined from the data in the corpus of documents. It is clear from Golden that data mining was used on a set of Earthquake data where relationships between a structured variable in the set (time of day) was related to those documents that contained California earthquake data (a sub-category of the corpus of earthquake data) using statistical measures. It would have been obvious to one of ordinary skill in the art at the time of invention to combine the teachings of Lewak and Golden as both documents discuss aspects of mining data sets. Adding Golden provides the benefit of using

statistical measures to determine whether or not relationships between earthquakes and time of day exist.

In regard to dependent Claim 2, Lewak teaches that *said algorithm comprises a keyword occurrence algorithm and wherein each of said categories comprises a category of text documents in which a particular keyword occurs* (Col. 8, lines 61-67; Col. 9, lines 1-4; 50-67; Col. 10, lines 1-10; categories (keywords) are further grouped based on sub-groupings, each sub-grouping containing similar documents, based on their categories (keywords)).

In regard to dependent Claim 25, Lewak fails to teach that *said calculating said probabilities comprises using a result of said counting said occurrences of said combination of said structured variables and said categories.* However, Goldman teaches this limitation (e.g. Pg. 17, Fig. 2). It would have been obvious to one of ordinary skill in the art at the time of invention to combine the teachings of Lewak and Golden as both documents discuss aspects of mining data sets. Adding Golden provides the benefit of using statistical measures to determine whether or not relationships between earthquakes and time of day exist.

In regard to dependent Claim 27, Lewak fails to explicitly teach *said providing said dictionary of keywords comprises generating said dictionary of keywords.* However, Goldman teaches this limitation (see Pg. 12, Sec. 2). It would have been obvious to one of ordinary skill in the art at the time of invention to combine the teachings of Lewak and Golden as both documents discuss aspects of mining data sets. Adding Golden provides the benefit of using statistical measures to determine whether or not relationships between earthquakes and time of day exist.

In regard to dependent Claim 28, Lewak fails to explicitly teach *determining whether said probability of occurrence of said combination of said structured variable and said category is below a predetermined value; and if said probability is below said predetermined value, designing said relationship as an interesting relationship.* However, Goldman teaches this limitation (see Pg. 16, Sec. 6.3 discusses statistical significance; those that managed to have statistical significance would validate or invalidate previous hypotheses such as that depicted on Pg. 17 and therefore would have been "interesting"). It would have been obvious to one of ordinary skill in the art at the time of invention to combine the teachings of Lewak and Golden as both documents discuss aspects of mining data sets. Adding Golden provides the benefit of using statistical measures to determine whether or not relationships between earthquakes and time of day exist. If they exist, they'd be crucial in predicting future events and thus interesting to both scientists and the public.

7. Claims 3-12, 14-16, 18-22, and 24 are rejected under 35 U.S.C. 103(a) as being unpatentable over Lewak in view of Goldman, and in further view of Goldszmidt et al. (hereinafter Goldszmidt, "*A Probabilistic Approach to Full-Text Document Clustering*", 1998, Technical Report ITAD-433-MS-98-044, SRI International).

In regard to dependent Claim 3, Lewak and Goldman fail to teach that *said algorithm comprises a clustering algorithm and wherein each of said categories comprises a category of said text documents containing a particular cluster.* However, Goldszmidt teaches both hierarchical agglomerative clustering as well as iterative clustering (such as K-means)(Pgs. 10-11, Sec. 3, 3.1, 3.2). It would have been obvious to one of ordinary skill in the art at the time of invention to combine the teachings of Lewak, Goldman, and Goldszmidt as all documents discuss aspects of grouping similar documents together. Adding Goldszmidt provides the benefit of using well known clustering techniques to group categories together and to compute probabilities used to measure the similarity between text-containing files (documents) to determine the similarity between them providing a gauge of how well the chosen categories (including structured variables) define the document content.

In regard to dependent Claim 4, Lewak and Goldman fail to teach that *said clustering algorithm comprises a k means algorithm.* However, Goldszmidt teaches both hierarchical agglomerative clustering as well as iterative clustering (such as K-means)(Pgs. 10-11, Sec. 3, 3.1, 3.2). It would have been obvious to one of ordinary skill in the art at the time of invention to combine the teachings of Lewak, Goldman, and Goldszmidt as all documents discuss aspects of grouping similar documents together. Adding Goldszmidt provides the benefit of using well known clustering techniques to group categories together and to compute probabilities used to measure the similarity between text-containing files (documents) to determine the similarity between them providing a gauge of how well the chosen categories (including structured variables) define the document content.

In regard to dependent Claim 5, Lewak teaches *said forming said categories comprises inputting a predetermined number of categories* (Col. 5, lines 28-31).

In regard to dependent Claim 6, Lewak and Goldman fail to teach that *said forming said categories comprises: generating a sparse matrix array containing a count of each of said keywords in each of said text documents.* However, generating a sparse matrix in this way is well known in the art and is typically a crucial part of most clustering algorithms.

In regard to dependent Claim 7, Lewak teaches that *said keywords comprise at least one of words and or phrases, which occur a predetermined number of times in, said text documents* (see Figs. 3-5 categories (keywords) can be words or phrases).

In regard to dependent Claim 8, Lewak and Goldman fail to teach *said calculating probabilities comprises using a Chi squared function*. However, Goldszmidt teaches using a Chi-Squared test as part of the analysis of clustering methods (p. 15, 3rd paragraph). It would have been obvious to one of ordinary skill in the art at the time of invention to combine the teachings of Lewak, Goldman, and Goldszmidt as all documents as all documents discuss aspects of grouping similar documents together. Adding Goldszmidt provides the benefit of using statistical measures to analyze clustering results.

In regard to dependent Claim 9, though Lewak fails to specifically teach that *said generating a dictionary of keywords comprises: first parsing text in said text document to identify and count occurrences of words; storing a predetermined number of frequently occurring words; second parsing text in said text documents to identify and count occurrences of phrases; and storing a predetermined number of frequently occurring phrases*, Lewak does either manually or automatically perform and provide a mechanism for compiling such a dictionary that involves viewing/parsing each of the uncategorized documents and determining, based on the subject matter (to include contemplating the number and meaning of descriptive term(s) or groups of terms) whether or not the term(s) or groups of terms are significant to describing the text document content. Thus, one of ordinary skill in the art at the time of invention would have considered such a method of compiling a list of keywords to be obvious based on well known and widely used techniques such as is contemplated by Lewak and as claimed.

In regard to dependent Claim 10, Lewak fails to teach that *said frequently occurring words and phrases are stored in a hash table.* However, it is typical to use hash tables as data structures, especially when the storage of vectors and matrices involved with clustering algorithms to enable their efficient storage and subsequent evaluation on a computer.

In regard to dependent Claim 11, Claim 11 contains subject matter that is similar to that found in Claims 1 (and similarly Claims 17 and 23) and 6, and is rejected along similar lines of reasoning.

In regard to dependent Claim 12, Lewak fails to teach that *said relationships comprise said combinations of structured variables and categories having a lowest probability of occurrence*. However, it is notoriously well known in the art that measures of whether or not two objects are grouped together or not depend on how closely or how distant characteristics of two objects are in comparison to one another. Those that are distant in terms of their similarities would translate to having a low probability of occurrence. Likewise, such similarity measures would also allow one to deduce how likely the clustering of two objects is the result of randomness.

In regard to independent Claim 14, Claim 14 reflects the method for identifying relationships between text documents and structured variables pertaining to said text documents as claimed in Claim 1 (and similarly Claims 17, and 23) and Claim 12, and is rejected along the same rationale.

In regard to dependent Claim 15 (and similarly dependent Claim 19), and dependent Claim 16 (and similarly dependent Claim 20), Lewak teaches that *said structured variables comprise predetermined time intervals* and *said predetermined time intervals comprise one of days, weeks, months and years* (see Figs. 3-5, category phrases involving time, date).

In regard to dependent Claim 18, Lewak and Goldman fail to teach *a memory for storing occurrences of said structured variables, categories and structured variable/category combinations and probabilities of occurrences of said structured variable/category combinations*. However, it would have been obvious to one of ordinary skill in the art at the time of invention to assume that such data would have to have been stored on some media such as memory, disk, or other computer storage, providing the benefit of ready access to the data for processing on a computer.

In regard to dependent Claim 21, Claim 21 contains subject matter similar to that found in Claim 14, and is rejected for similar reasons.

In regard to dependent Claim 22, Lewak and Goldman fail to teach that *said relationships comprise statistically significant relationships.* However, Goldszmidt teaches both hierarchical agglomerative clustering as well as iterative clustering (such as K-means)(Pgs. 10-11, Sec. 3, 3.1, 3.2). The determination of similarity is at the heart of most clustering algorithms because it is that measure that allows those algorithms to group similar documents together. Even if done manually, as in the teaching of Lewak, a human being of ordinary skill would have been able to produce groupings of documents that would have been statistically significant. It would have been obvious to one of ordinary skill in the art at the time of invention to combine the teachings of Lewak, Goldman, and Goldszmidt as all documents as all documents discuss aspects of grouping similar documents together. Adding Goldszmidt provides the benefit of using well known clustering techniques to group categories together and to compute probabilities used to measure the similarity between text-containing files (documents) to determine the similarity between them providing a gauge of how well the chosen categories (including structured variables) define the document content.

In regard to dependent Claim 24, Lewak teaches that *said structured variables comprise structured data* see Figs. 3-5, category phrases involving time, date).

Response to Arguments

8. Applicant's arguments, see amendment, filed 02/21/2006, with respect to the rejection(s) of claim(s) 1-12, and 14-24 under Lewak in view of Goldszmidt have been fully considered and are persuasive. Therefore, the rejection has been withdrawn. However, upon further consideration, a new ground(s) of rejection is made in view of Goldman et al.

9. Specifically, Applicant argues that the prior art of Lewak and Goldszmidt either alone or in combination fail to teach the amended limitation of *identifying a relationship between a structured variable (of the structured variables) and text documents included in a category (of the categories) based on a probability of occurrence of a combination of the structured variable and the category*. The Examiner would tend to agree and withdraws the rejection. However, upon further searching, the Examiner now adds the prior art of Goldman et al., which, in combination with the previous prior art teaches the amended limitation. Goldman performs knowledge discovery on a text database of Earthquake data looking for correlations between earthquakes (in a sub-category of California, as an example) and time of day (structured variable) of the Earthquake's occurrence using statistical measures to determine significance of the combination of the two.

Conclusion

10. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

11. A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

12. Any inquiry concerning this communication or earlier communications from the examiner should be directed to James H. Blackwell whose telephone number is 571-272-4089. The examiner can normally be reached on Mon-Fri.

13. If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Heather R. Herndon can be reached on 571-272-4136. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

14. Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

James H. Blackwell
05/10/2006

William L. Bashore
WILLIAM BASHORE
PRIMARY EXAMINER
5/11/2006